A Study about Collecting Preference Information System for Creating Outing Plan

Kazuya Yamashita† Takeshi Iwamoto‡ and Michito Matsumoto†

† ‡ 5180 Kurokawa Imizu, Toyama Prefectural University, Toyama, Japan
E-mail: † 828kazuya@yamamoto@puc.pu-toyama.ac.jp, ‡ {iwamoto, michito}@pu-toyama.ac.jp

Abstract Today, there are various planning systems that create outing plans for trips and drives, and support for tourists. The existing systems are designed to use user preference information to plan suitable trip plans for the user. However, there are some problems with the current systems. The systems require the user to select a lot of information by a questionnaire. Moreover, this still does not provide enough information to understand the user preference. The system we have designed resolves some of these problems by collecting user preference information by using GPS data from daily life. The data is used to estimate visited facilities and daily life patterns. By collecting such a user preference information, we can realize the planning system which reduce user’s burden. In this paper, we proposed method to collect visited facilities information from GPS data of daily life.

Keyword Mobile History, Life Log, Preference Information, Clustering

1. INTRODUCTION
   Today, there are various planning systems for tourists when tourists go out. These systems create outing plans for trips and drives, and systems support for tourists. The existing systems are designed to use user preference information to create suitable trip plans for the user. However, there are some problems with the current systems. The existing systems usually require the user to select a lot of information although sometimes this still does not provide enough information to understand the user preference. So in this study, we resolve some of these problems by collecting user mobile history information from smartphones using GPS and collecting preference information from mobile history information. By collecting the user preference information for the planning system, problems can be resolved.

2. PREFERENCE INFORMATION
   In some existing studies, preference information is defined by retrieved music history information and information about extracted personal preference on ahead. In this research, we define preference information as users’ prefer place where users usually visit such as a shop, restaurant and so on. If this preference information are collected and used, planning system easily can create outing plan that comprehend user preference. For example, planning system can give users a suggestion of an appropriate sightseeing spot and sights by using feature of facility information where the user visits every day. In addition, when planning system keeps label on user favorite dish and restaurant, system can reflect user preference information to outing plan.

3. SYSTEM SUMMARY
   In this study, the system collects user mobile history information by using GPS data on a smartphone. The system focus on finding visited facilities and they could be mainly indoor. When system collects GPS in the indoor, GPS have significant errors and it causes problems. However, when we collected GPS by using system in fact, we detected that GPS have not random but unique errors. We found that the error can be used as feature of each facility. Therefore, we use GPS errors to find visited facilities. The method collects GPS data and divides into several clusters. And system labels the clusters as facility attribute information. In this way, system deal with visited facility attribute information as preference information. Fig.1 shows summary of proposed system. First, collecting preference information system estimates visited facility by using clustered GPS data. Second, system labels information obtains from actually existing facility to estimated visited facility. The information is assumed to include, attribute information such as name, categories, prices of goods and so on. They can be used as user preference information. Third, system can reflect user preference information to create outing plan on planning
system by using collected preference information stored at the database.

![Diagram](Image)

**Fig. 1. System Summary**

### 4. DESIGN AND IMPLEMENTATION

In this chapter, we explain collecting user preference information method by using clustering. First, system uses clustering and labeling. Thus, we need to confirm accuracy of collecting mobile history information. So in this chapter, we make an evaluation on clustering method and labeling method. In evaluation of clustering method, we make an evaluation whether number of derived clusters by using this system and number of visited facilities are correct or not. And, we also make an evaluation of time-series clustering. We use 19 days of collected GPS at 30 second intervals on a smartphone for evaluation of clustering. Table 1 shows summary of data. And in evaluation of labeling method, we make an evaluation whether system labels accurately facility attribute information.

#### Table 1. Data Summary

<table>
<thead>
<tr>
<th>Number of subjects</th>
<th>2 subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collected GPS (days)</td>
<td>19 days</td>
</tr>
<tr>
<td>Visited facilities</td>
<td>52 places</td>
</tr>
<tr>
<td>Average visited time</td>
<td>24.7 min</td>
</tr>
<tr>
<td>Maximum visited time</td>
<td>153.5 min</td>
</tr>
<tr>
<td>Minimum visited time</td>
<td>5 min</td>
</tr>
</tbody>
</table>

In evaluation of clustering method, we use two parameters that visited time of one cluster and cluster size to be judged the same cluster. We define visited time that assumed visited time of one cluster, and we set visited time to 3, 5, 10, 15, 20 minute for evaluation. Also cluster size is set to 1 meter from 1 to 100 meter. Fig. 2 shows relationship between cluster size and percentage of correct on each visited time. As a result, in case of visited time are 3 and 5 minute, we got high percentage of accuracy under certain cluster size. But in case of visited time are 10, 15, 20 minute, we did not get high percentage of accuracy. We considered 2 reasons. The first cause is that subjects did not visit facilities for a long time. The second cause is that system extra derived clusters in small cluster size. Therefore, we did not get high percentage of correct.

![Graph](Image)

**Fig. 2. Relationship between cluster size and percentage of correct on each visited time**

We also make an evaluation of time-series clustering. To do this evaluation, we use determined time to be judged...
the same cluster in addition to clustering parameters. Determined time is threshold to decide whether single visit or not. If the timestamp of GPS data are beyond determined time, system can judge these data are occurred by several visit of the user. We set determined time to 1, 2 hours for an evaluation if time-series clustering. Fig.3-1 and Fig.3-2 show relationship between determined time and cluster size. From Fig.3-1 and Fig.3-2, in case of short visited time, we also got high percentage of accuracy under certain cluster size. We consider that short determined time causes that system derived into too small clusters even if user visited the facility at once.

We set determined time to 1, 2 hours for a valuation if time-series clustering. Fig.3-1 and Fig.3-2 show relationship between determined time and cluster size. From Fig.3-1 and Fig.3-2, in case of short visited time, we also got high percentage of accuracy under certain cluster size. We consider that short determined time causes that system derived into too small clusters even if user visited the facility at once.

We considered 2 causes of this result. The first cause is that system mistakenly labeled as different facility attribute information from GPS error. The second cause is that foursquare API did not have visited facility information. In this study, we focus second cause, and resolve this cause by using labeling range. If system is set labeling range, system does not label the information placed out of the range. Fig.5 shows result of evaluation when system uses labeling range. From Fig.5, we learned that result is slightly changed in case that range is set from 140 to 220 meter. But, we could not dramatically improve percentage of accuracy by only using labeling range. So, we need to use other pieces to improve this result.
6. CONCLUSION

In this study, we collected user preference information to reduce user’s burden such as complicated questionnaire. By clustering and labeling GPS that system collected user’s preference from user daily life. In collecting preference information method by using clustering, the system was able to collect suitable user mobile history information under certain conditions. In addition, system showed the usefulness of time-series clustering. And system collected user preference information when system labeled collected mobile history information as facility attribute information. In addition, labeling range had an influence on percentage of correct answer. On the basis of past achievements, we consider that system can label accurately by using GPS error.

In challenges for the future, we need to collect a lot of GPS data to obtain better result. If we collect more GPS data, we might analyze better clustering parameters conditions and improve labeling accuracy. This study, we have not yet implemented system for creating suitable plan for user. Therefore, we need to implement planning system that adapts user preference information.

References

[1] foursquare for Developers
https://developer.foursquare.com/


